```
OM Of: US-09-494-297-2 to: EST:*    out_format : pfs
Date: Jun 6, 2001 11:26 PM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL -frame+_p2n.model -DEV-xlp
-Q-/cgn1_1/USPTO_spool/US094j4297/tunat_06062001_115734_12297/app_query.fasta_1.825
-DB-EST -QFMT-fastap -SUFFIX-rst -GAPOP-12.000 -GAPEXT-4.000
-MINMATCH-0.100 -LOOPCL-0.000 -QGAPOP-4.500
-QGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500 -FGAPOP-6.000
-FGAPEXT-7.000 -XGAPOP-10.000 -XGAPEXT-0.500 -DELOP-6.000
-DELEXT-7.000 -START-1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST-45 -DOCALIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0
-ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-ext -MINLEN-0
-MAXLEN-200000000 -USER-US094j4297_@CGN1_1_3891 -NCPU-6
-ICPU-3 -LONGLOG -NO_XLPXV -WAIT -THREADS-1

Search information block:
Query: US-09-494-297-2
Query length: 757
Database: EST:*
Database sequences: 9623517
Database length: 73081774
Search time (sec): 1597.010000
```

```
seq_name: gb_est80:BE919552

seq_documentation_block:
LOCUS       BE919552      626 bp    mRNA            EST       02-OCT-2000
DEFINITION  EST423321 potato leaves and petioles Solanum tuberosum cDNA clone
            cSTB1D12 5' sequence, mRNA sequence.
ACCESSION   BE919552
VERSION     BE919552.1  GI:10445628
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
            Utterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning
            ,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
  TITLE     Generation of ESTs from potato leaves and petioles
  JOURNAL   Unpublished (2000)
  COMMENT   Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone request: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com.
FEATURES             Location/Qualifiers
     source          1..626
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone_lib="potato leaves and petioles"
                     /clone="cSTB1D12"
                     /tissue_type="leaflets and petioles"
                     /dev_stage="8 weeks old plants"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site.1: EcoR1; Site.2:
                     XhoI; Tissue was supplied by Dr. Fry (Cornell University).
                     leaflets and petioles were isolated from 8 week old
                     greenhouse grown plants. The plants were watered and
                     fertilized freely. The tissue was immediately frozen in
                     liquid nitrogen."
BASE COUNT     172 a    108 c    168 g    178 t
ORIGIN

alignment_scores:
        Quality:    108.00          Length:   248
        Ratio:      0.878           Gaps:      12
        Percent Similarity: 49.597  Percent Identity: 24.194

Align seq 1/1   to: BE919552   from: 1   to: 626

alignment_block:
US-09-494-297-2 x BE919552
```

```
193 AATGGAAAAGAAAAATTGGAGGAAAAGATGTTAGAGAGHAGTCAAAAAGA 242
        |||        |||::::::            |||
509 laAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuAsp 525

243 TAGCT...CAAAATAAGAAACATTATGAACAAACTTGAAAAAAGAATTA... 286
        :::          :::||||||        :::  ||||||
526 LysAspLysLeuLysAspTyrHisGlyPheGlyAspMetAsnAspSerTh 542

287 AAAGAAAAAACAAGAAGAAATAC......TTTGGAGAATTAAACACATCAA. 328
        ||||||:::::::||              |||||||:::::::::.:::
542 rLeuAlaValAlaLysIleLeuValGluTyrAlaGlnAspSerAsnPro 559

329 .ATTGGAGTAGAAAATATTATTTATTTAGAATGTAGTGAAAGAATAAATAAAG 377
        ::::||  ||::::||||:||||||::::||
576 SerLeuIleGlyThrGlnTrpHisProGluAspLeuValAspIleIleAr 592

378 AAAGA......GAAATTGAAATGTTT............... 397
        :::          :::||||||:::.||||
559 roGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLysTyrGln 575

398 ..........AACATTCAACAAGAGAAAATTGATGTACTTAAGAACGAATTAA 438
        ||        |||||::::||||:::|||:::        ||||||:::
592 gMetGluAspLysLysGluValIleProValThrHisAsnLeuThrLeuA 609

439 AATAGAAGAAGAAGAGAAT........AGTTCAACAGAAAAATGGAATTAG 482
        ||||:::::::::||||:||          ::::||||||:::|||||:::
609 rGlyLysThrValThrGlyLeuAlaGlyTyrAspArgThrLysAspPheHisPhe 625

483 AAAAAGAGAAATTTAT............GATCTTAAAGAGAGAGATTCAACA 520
        :::||||           ||            ||||||:::|||
642 sThrAspLysThrAsnLeuGluPheLysAspGlyLysAlaThrIleAsnL 659

521 GAAATT.........AATGAAAAAGGAGAGTGAAAATTAAACAAAAGCTTGA 561
        ||||||            ||||||:::||||||::::||||||||
626 GluIleGluLeuLysGlnGluLeuLeuSerGlnThrValVal 642

562 AAGAAAATAATACAGAAATTGAAAATAAAGTGGAGAATTACGCTAGTGTAT 611
        ::::||||||::::||||||:::::||||||:::||::::||||:::
659 euLysHisGlyGluSerLeuThrLeuGlnIleGlyLeuProGluGlyTyrSer 675

612 TAAGAGAAATGGAAAACTAAAATGAATGGTTACAGATAAAAAAGAAAGAGA 661
        :::          |||||||:::          |||||:::||
676 Tyr..........LeuValLysGluThrAspSerGluLysGlyLyrLy 687

662 TATGAAGAAGTTGAAAATCTTATTAAAACAAAATCAAAAAGA...... 703
        ::::||||||:::||||||                ::::|||
687 sValLysValAsnSerGlnGluValAlaAlaAsnAlaThrValSerLysThrG 704

704 CAAAGTCATCGCACAAAGTATGTATACTAAAAAGTTGC 737
        ||||::::||||||||:::|||:::|||::::||||||
704 lyIleThrSerAspGluThrLeuGlnAsnAsnLysGluProVal 720

738 AAGTTGGATTAGATGAAAACGTTCGAAGTGCTTAGAATGGCAACAAATAGA 787
        ::::
721 ValProThrGlyValAspGlnLys 728

788 TTACCTAATTAAAGACAAATAGAAAG 811
```

```
503 ThrGluThrGlnLeuGluArgAlaAlaThrGlnLeuAlaIleAlaIleTyrTyrPheTh 519
        ||||||:::||||||:::|||         :::       |||
  9 ACTGAAAGCTCAGATAAAGGACAATTGTGGAGTGGTTATTGCTTACCATGG 58

519 rAspSerAlaGluLeuAspLysAspLysLeuLysAspTyr......TyrHisG 534
        |||:::        ||||||:::||||||            |||:::
 59 AGACTCAACACAGGTTTAAGTACTGACAGTTTGGCGTGATCGTGGGTATCCCG 108

534 lyPheGlyAspMetAsnAspSerThrLeuAlaValAlaLysIleLeuVal 550
        ||||||        :::|||::::||||||:::         |||
109 GACGCAGCTTCACTTGGTGAT............GCAGTTGTGGTATGGGCGTGTC 149

550 GluTyrAlaGlnAspSerAsnProGlnLeuThrAspLeuAspPhe.. 566
        :::|||||||:::         ::::|||||||||
151 ..........ATAACTTCTAAAGAATTTCTT 175

567 ......PheIleProAsnAsnAsnLysTyrGlnSerLeuIleGlyThrG 581
        |||::::||||||:::||||||:::||||||:::|||
176 GTTTTGGTTTCGGTCCCCACACACGCAAAAGAAAAAAGTGGGGTGGTGCA 225

581 lnTrpHisProGluAspLeuValAspIleIleArgMetGluAspLysLys 597
        :::||||||||||||:::         |||
226 AGCATCACCTGAAGACAAGGAATGACCGACAGAGAATG............ 263

598 GluValIleProValThrHisAsnLeuThrLeuArgGlyLysThrVal 614
        |||:::       |||||||:::        ||||||
264 ..........CATCCACGTTCTTCCTTCAAGGCATTTCTGA 295

614 yLeuAlaGlyAspArgThrLysAspPhe...HisPheGluIleGluLeu 630
        ::::||||||:::||||||:::       |||:::||
296 AGTTTGTTAAAGTCGGAGCTCGCCCATGCGAAAATGCTGAAATGGATGCAA 345

630 ysAsnAsnLysGlnGluLeuLeuSerGlnThrValValLysThrAspLysThr 646
```

seq_name: gb_est38:AV603854

```
346 TCCACTCTTTGCAGCTPAATTCTCCGAGATTCATTTAAG............. 383
    :::::                :::::             |||
647 AsnLeuGluPheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGl 663
384 GATGCTGAGGCAAGTAATTCTAAGGCTTATTGTTGGCATCGTCATCTTGGGA 433
    ||||||||||||||||||       ||||||     ||||||
434 A..ATGGAGTTGCAAGGGATAGATGAA...CTGAGTTCTGTTGCCAGG 477
663 uSerLeuThrLeuGlnLeuProGlnGlyTyrSerTyrLeuValLysG 680
    |||||||||||||                 |||||||||||||
478 A..ATGGTTAGATTGAATCGAAAACTGCA 503
680 luThrAspSerGluGlyTyrLysValLysAsnSerGlnGluValAla 696
    ||||||::||||||||                ||||::||
504 ACAGCT............. 509
697 AsnAlaThrValSerLysThrGlyIleThrSerAspGluThrLeuAlaPh 713
    :::||                   ||||||||::
510 .........CCCATATTGCTGTTGAATGTCGAAGGTCGCATAA 543
713 eGluAsnAsnLysGluProValProThrGlyValAlaAspGlnLysIleA 730
    |||||                      :::||||||||
544 ATGGTTGGAATGCAAAGGTCGCTGCTGAATTGACAGGTTTATCAATT 587
730 snGlyTyrLeuAlaLeuIleVal...IleAlaGlyIleSerLeu 743
    |||:||||||||||||||||
544 ATGGTTGGAATGCAAAGGTCGCTGCTGAATTGACAGGTTTATCAATT 587
```

seq_documentation_block:
LOCUS   AV603854   584 bp   mRNA   EST   30-AUG-2000
DEFINITION AV603854 Bos taurus kidney fetus Bos taurus cDNA clone E1KI019C04
            5', mRNA sequence.
ACCESSION   AV603854
VERSION   AV603854.1   GI:9734227
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 584)
  AUTHORS   Sugimoto,Y., Hirotsune,S., Takasuga,A., Itoh,R., Jitohzono,A. and
            Suzuki,H.
  TITLE   bovine cDNA sequencing
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            This clone was obtained from a polyA-deleted cDNA library.
            Single pass sequencing.
FEATURES   Location/Qualifiers
     source   1..584
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1KI019C04"
                     /clone_lib="Bos taurus kidney fetus"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
                     was deleted from a NotI site."
BASE COUNT   196 a   148 c   131 g   109 t
ORIGIN

alignment_scores:
    Quality: 107.50   Length: 239
    Ratio: 1.097   Gaps: 8
```

Percent Similarity: 41.004        Percent Identity: 24.268

alignment_block:
US-09-494-297-2 x AV603854

Align seq 1/1 to: AV603854   from: 1   to: 584

US-09-494-297-2:   .          AV603854:   from: 1   to: 584

```
264 AsnLeuLeuSerGlyGlyLeuValProThrLysProProThrProGlyAs 280
    |||||||||||||  :::                ||||||||||||
39  AATCTCCTGGAT......TTACCACCCCCCATCTCCTCCCAAACCAAAAAC 82
280 pProProMetProProAsn...............GlnProGlnThrThrS 292
    :::||||||||||:::               :::||:::
83  CAATTGTCTTTACCCTCCCAACTGGAAGACACGCCCCGAGACCCCGAGGAAAG 131
292 erValLeuIleArgLysTyrAlaIleGlyAspTyrSerLysLeuLeuGlu 308
    :::||            ||||||||:::
132 ...ATTTACTACTACCACCATGTGGTCCACCAGGCAGACTCAGTGGGAT 173
309 GlyAlaThrLeuGlnLeuThrGlyAspAspAsnValAsnSerPheGlnAlaAr 325
    ||||||       :::|||||||::::::
174 CCTCCCTACTTGGGAAAGCCCCAGGAGGATGAGATGCCCAGC........... 209
325 gValPheSerSerAsnAspLeuGlyLeuArgIleGluLeuSerAspGlyThr 342
    :::              ||||||||||||
210 ...CTTGACCATGAAAGCTGAAGATGGACTTGGGAA 240
342 hrTyrThrLeuThrGluLeuAsnSerProAlaGlyTyrSerIleAlaGlu 358
    ||  |||                    :::
241 CCCCAACATATGAT................................... 254
359 ProIleThrPheLysValGlnUlaAlaGlyLysValTyrThrIleAlaAspGl 375
    |||                              :::
254 ...........................................GACACCTCCAGTGAGCTGGCTAAGAAAAGCAAA 332
375 yLysGlnIleLeuGluAsnProAsnLysGlyGluIleValGluProTyrSerValG 392
    ||||||||||   :::              ||||::::
255 ...GAAAATCCCATGAGACCCCAAAAAAGCCTAAGACAGCAG 294
392 luAlaTyrAsnAspPheGluGlyPheSerValLeuThrThrGlnAsnTyr 408
    ||  |||                         :::
295 AAGCA.......GACACCTCCAGTGAGCTGGCTAAGAAAAGCAAA 332
409 AlaLysPheTyrTyrAlaLysAsnGlySerSerGlnValValTyr 425
    :::|||||||||                  ||||::
333 GAAGTATTC...........AGAAAAGAGATGTCCCAGTTTATCGTCCA 370
425 rCysPheAsnAlaAspLeuLysSerProProAspSerGluAspGlyGlyl 442
    |||:::             ||||||||||
371 GTGCCTGAACCCTTACCGGAA......CCTGACTGCAAAGTGGGAAGA 414
442 ysThrMetThrProAspPheThrThrGlyGluValLysTyrThrHisIle 458
    |||                    ||:::
415 TTACCACAACCGGAGAGACTTCAAACACCTAGCTCCGCAAGCTGACTCATGGC 464
459 AlaGlyArgAspLeuPheLysTyrThrValLysProArgAspThrAspPr 475
    :::||||||||               ||||:::
459 GTTATGAATAAGGAGCTGAAGTACTGTAAGAACCCCGAGGACCTGGAGTG 514
475 oAspThrPheLeuGluHisIleLeuLysValIleGluLysGlyTyrArg 492
    :::||||||||                 ||||:::
465 CAATGAGAATGTGAAACACAAAACCAAGGAGTACATTAAGAAGTACATGC 564
492 luLysGlyGlnAlaIle 497
    |||            |||
515 ...........................................CAATGAGAATGTGAAACACAAAACCAAGGAGTACATTAAGAAGTACATGC 564
565 AGAAGTTTGGGGCTGTT 581
```

seq_name: gb_gss23:AZ527959

seq_documentation_block:
LOCUS   AZ527959   915 bp   DNA   GSS   03-NOV-2000
DEFINITION ENTDH31TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.

```
ACCESSION   AZ527959
VERSION     AZ527959.1  GI:11080143
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 777.
FEATURES            Location/Qualifiers
     source         1..915
                    /organism="Entamoeba histolytica"
                    /strain="HM1:IMSS"
                    /db_xref="taxon:5759"
                    /clone_lib="Entamoeba histolytica sheared DNA"
                    /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                    Institute for Genomic Research (TIGR), Rockville, MD.
                    Genomic DNA isolated from broth cultures of E. histolytica
                    using a method described by Clark and Diamond (Clark,
                    C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a
                    method for isolate identification. Exp. Parasitol.
                    77:450.). The DNA was mechanically sheared to give a
                    tight size distribution (~2 kb). The v + l method used for
                    the library construction is described in detail in Smith,
                    H.O. and Venter, J.C. (Making small insert libraries for
                    whole genome shotgun sequencing projects. In Genome
                    Sequencing: A Practical Approach, eds. M. Vaudin and B.
                    Barell, Oxford University Press, 1999)."
BASE COUNT      429 a     98 c    141 g    247 t
ORIGIN
```

```
alignment_scores:
     Quality:   102.50               Length:   343
     Ratio:      0.674               Gaps:      12
Percent Similarity:  44.315    Percent Identity:  17.784

alignment_block:
US-09-494-297-2 x AZ527959

Align seq 1/1  to: AZ527959       from: 1 to: 915
```

```
417  LysAsnGlySerSerGlnValValTyrCysPhe...............AsnAl  429
      ||||||       |||||||||||||||          :.:::
17   AAAAATTAAATGTCAAGAGTAGTAGTTAGATTATTTATTTAAATGGACAATG  66

429  aAspLeuLysSerProProAspSerGluAspGlyGlyLysThrMetThrP  446
      :::::|||        |||||| |||||||       ||||||||
67   CGTTGCAACTAAACCACTAAATTCACAAGACAAATTGAAAACAGCA....  112

446  roAspPheThrThrGlyGluValLysThrThrHisIleAlaGlyArgAsp  462
                                                          112

463  LeuPheLysThrTyrThrValLysProArgAspThrAspThrPheLe  479
      :::|||      ::|||  ||  :  |||  |  ||||
113  ......AGAGAGAAAATAAAAGAAAAGAGATGGTCTGACTCTCAACACTTTTT  156
```

```
479  uLysHisIleLysLysValIleGluGluLysGlyTyrArgGluLysGlyGlnA  496
      :::|||||:.:::|||::                   |||
157  AACAAAAGAAGGAGGATAATAATTGATAAGAATGATGAAGAG.........  196

496  laIleGluThrTyrSerGlyValGluThrGluThrGlnLeuArgAlaAlaThrGln  512
                                                              196

513  LeuAlaIleThrTyrThrPheThrAspSerAlaGluLeuAspLysAspLysSle  529
      ...TGTTTTACTGTGAGGAGATGTAATTGTAGAAAAAGAGT  234

529  uLysAspTyrHisGlyPheGlyAspMetAsnAspSerThrLeuAlaVal.  545
      |||||    |||:::       |||     :::    |||
235  AATAAAATATTAAAGGAATAGAAGAGGAATAAAAGAGAATAAAGATAA  284

546  .AlaLysIleLeuValGluTyrAlaGlnAsnProSerAsnProGlnLeu  561
      ::||||||   |||||          |||    |||   |||
285  ATGATAAGATAATTAACTACTATTGAAATAGATAATTAAAACACCA...TTA  331

562  ThrAspLeuAsp........                    PhePh  567
      ::|||:|||                             ::::
332  AGTGATGTTAGAAAAATAGTTCAAAACAATTCCAGATTCAGCATATTTCTA  381

567  eIleProAsnAsnAsnLysTyrGlnSerLeuIleGlyThrGlnTrpHisAsp  584
      ::::|||::|||:|||     :::
382  TACACTTGACAATGATAAAATGAAAGAAATACTGAAGAAGAAATTTATAG  431

584  roGluAspSerProLeuVal......AspIleIleArgMetGluAspLysLysGlu  598
      |||||:::|||::::::        |||   |||  :::::||||||||
432  TAGAAGACAATAATAAATAATGATGAAAATTATTCTTAAAGAAGAAAAACAGA  481

599  ValIleProValThrHisIleAsnLeuThrLeuArgLysThrValThrGlyIle  615
      |||||||                :::::
482  AATATACCAGAAATCAATGATATAACAAATTGGAATAATGTTTGAATGGAA  531

615  uAlaGlyAspArgThrLysAspPheHisPheGluIleGluLeuLysAsnA  632
      :::      |||:|||:::       |||::|||:|||:::
532  ACCAAAGATAAAA...AAGCAATTCAACAAGAATATTTCTTCAGATC  578

632  snLysGlyGluLeuLeuSerGlnThrValLysThrAspLysThrAsnLeu  648
      :::::|||::::|||:::       :::
579  TTAGAAAAAGAAATTGAAAATGAA......                      601

649  GluPheLysAspGlyLysAlaThrIleAsnLeuLeuLysHisGlyGluSerIle  665
      ::::::
601  ......................................................  601

665  uThrLeuuGlnLeuProGluGlyTyrSerTyrLeuValLysGluThrA  682
      |||  :::                                |
602  ......AAAGAAATTCCAAAAGACTTTCTTT.................GAAG  632

682  spSerGluGlyTyrLysValLysValAsnSerGlnGluValAlaAsnAla  698
      |||||::|||||||||||:::|||||||||||:::
633  ATCAAGATGGATTTAAAATATCAAATTAATGATGAACA.........  670

699  ThrValSerLysThrGlyIleThrSerAspGluThrLeuAlaPheGluAs  715
      :::::||||       |||||:::
671  TCATTAAAACTAACATCTATTTTGCACGATAATAAAACTCAATATAAAAAAC  720

715  nAsnLysGluProValValProThrThrGlyVal......       AspG  727
      |||  :::|||                                  |||:
721  AGAAGTTACTGATTACCCTCCTTTATAATTATTTGAAGTTAGTCTTTTAAGATG  770

727  lnLysIleAsnGlyTyrLeuAlaLeuIle  736
      ::|||            |||:||
771  AGAAAGAAAATCACCAGGAATCAATGATA  799
```

```
seq_name: gb_est49:AW654454

seq_documentation_block: AW654454
LOCUS       AW654454     508 bp     mRNA       EST       14-JUL-2000
DEFINITION  103924 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
```

```
ACCESSION   AW654454
VERSION     AW654454.1  GI:7420280
KEYWORDS    EST.
SOURCE      cow.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 508)
AUTHORS     Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
TITLE       Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
JOURNAL     Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAAGCATATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 88  row: B  column: 3
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES    Location/Qualifiers
     source  1..508
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 1BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
            library made from pooled tissue from lymph node, ovary,
            fat, hypothalamus, and pituitary."
BASE COUNT  169 a   117 c   135 g    87 t
ORIGIN
```

```
alignment_scores:
   Quality:  101.50              Length:  129
   Ratio:      1.538             Gaps:      4
   Percent Similarity: 51.163    Percent Identity: 31.008

alignment_block:
US-09-494-297-2 x AW654454

Align seq 1/1  to: AW654454   from: 1  to: 508    ..

379 GluAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAs 395
    ||||||||||| :::||||| :::::||||||||||||
379 GluAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAs 395
111 GAAAATCCATGAAGAGACTTCAAAAAAGCCTAAGACAGACA....... 155

395 nAspPheGluGluPheSerValLeuThrThrGlnAsnTyrAlaLysPheT 412
    :::|||
156 ......GACACCTCCAGTGAGCTGGCTAAGAAAAGCAAAGAAGCTATTC. 197

412 yrThrTyrAlaLysAsnLysAsnGlySerSerGlnValValTyrCysPheAsn 428
    ::|||||:||
198 .........AGAAAAGAGATGTCCCAGTTATCGTCCCAGTGCCTGAAC 236

429 AlaAspLeuLysSerProProAspSerGluAspGlyGlyLysThrMetTh 445
    |||||||||
237 CCTTACCGGAAA.......CCTGACTGCAAAGTGGGAAGAATTACCACAC 280

445 rProAspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgA 462
    ||| |||||| :::
281 CGAAGAGACTTCAAAACACCTAGCTCGCCAAGCTGACTCATGGCGTTATGAAATA 330
```

```
seq_documentation_block:
LOCUS       AL037842    570 bp    mRNA    EST    29-FEB-2000
DEFINITION  DKFZp564H077_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone
            DKFZp564H077 5', mRNA sequence.
ACCESSION   AL037842
VERSION     AL037842.1  GI:5407173
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 570)
AUTHORS     Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
TITLE       EST (Bloecker, et al.)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Bloecker H
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp564H077) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES    Location/Qualifiers
     source  1..570
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="DKFZp564H077"
            /clone="DKFZp564H077"
            /tissue_type="brain"
            /dev_stage="fetal"
            /lab_host="xl-2blue"
            /note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
BASE COUNT  190 a   118 c   152 g   110 t
ORIGIN
```

```
seq_name: gb_est28:AL037842

462 spLeuPheLysTyrThrValLysProArgAspThrAspProAspThrPhe 478
    :::|||||| |||||||| :::
331 AGGAGCTGAAGTACTGTAAGAACCCGAGGACCTGAGTGCAATGAGAAT 380

478 LeuLysHisIleLysLysValIleGluLysGlyTyrArgGluLysGlyGl 495
    :::|||||| ||| ::|||
381 GTGAAACACAAAACCAAGGAGTACATTAAGAAGTACATGAGAAGTTTGG 430

495 nAlaIleGluTyrSerGlyLeuThrGluThrGlnLeu 507
    ||| :::|||:||||
431 GGCTGTT...TACAAACCAAAGAGGACACACTGAGTTA 464
```

```
alignment_scores:
   Quality:  101.50              Length:  129
   Ratio:      1.538             Gaps:      4
   Percent Similarity: 51.163    Percent Identity: 31.008

alignment_block:
US-09-494-297-2 x AL037842

Align seq 1/1  to: AL037842   from: 1  to: 570    ..

379 GluAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAs 395
    ||||||||||| :::||||| :::::||||||||||||
379 GluAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAs 395
130 GAAAACCCATGAAGGCCTCGAAAAAGCCCAAGACAGCAGAAGCA...... 174

395 nAspPheGluGluPheSerValLeuThrThrGlnAsnTyrAlaLysPheT 412
    |||:::
```

Left column:

```
175 ........GACACCTCCAGTGACTAGCAAAGAAAAGCAAAAGAAGAGTATC. 216
          :::||||||
412 yrTyrAlaLysAsnGlySerSerGlnValValTyrCysPheAsn 428
          :::||:|||
217 .........AGAAAAGAGATGTCCCAGTTCATCGTACGAGTGCCTGAAC 255

429 AlaAspLeuLysSerProAspProAspSerGluAspGlyGlyLysThrMetTh 445
          |||||||||
256 CCTTACCGGAAA......... CCTGACTGCAAAGTGGGAAGAATTACCACAC 299

445 rProAspPheThrThrGlyGluValValLysTyrThrHisIleAlaGlyArgA 462
          | ||||||
300 TGAAGACTTTAAACATCTGGCTCGCAAGCTGACTCACGGTGTTATGAATA 349

462 spLeuPheLysTyrThrValLysProArgAspThrAspProAspThrPhe 478
          :::||||||
350 AGGAGCTGAAGTACTGTTAAGAAATCCTGGAGACCTGGAGTGCAATGAGAAT 399

479 LeuLysHisIleLeuLysLysValIleGluLysGlyTyrArgGluLysGlyGl 495
          ::|||||||
400 GTGAAACACAAAACCAAGGAGTACAATTAAGAAGTACATGCAGAGAAGTTGG 449

495 nAlaIleGluTyrSerGlyLeuThrGluThrGlnLeu 507
          |||:|||
450 GGCTGTT...TACAAACCCAAAGAGGACACTGAATTA 483
```

seq_name: gb_est28:AL037979

seq_documentation_block:

LOCUS	AL037979 673 bp mRNA EST 29-FEB-2000
DEFINITION	DKFZp564P077_r1 564 (synonym: hfbr2) Homo sapiens cDNA clone
	DKFZp564P077 5', mRNA sequence.
ACCESSION	AL037979
VERSION	AL037979.1 GI:5407291
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 673)
AUTHORS	Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Gassenhuber,J. and
	Wiemann,S.
TITLE	EST (Bloecker, et al.)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Bloecker H
	MIPS
	An Klopferspitz 18a D-82152 Martinsried, Germany
	This is the 5' sequence of the clone insert
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
	Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
	sequenced by GBF (National Research Centre for Biotechnology Ltd.,
	Braunschweig/Germany) within the cDNA sequencing consortium of the
	German Genome Project.
	sl sequence also available.
	This clone (DKFZp564P077) is available at the RZPD in Berlin.
	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
	Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES	Location/Qualifiers
source	1..673
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="DKFZp564P077"
	/clone="DKFZp564P077"
	/tissue_type="brain"
	/dev_stage="fetal"
	/lab_host="xl-2blue"
	/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

BASE COUNT   222 a   147 c   171 g   133 t   Length: 129
ORIGIN

alignment_scores:
    Quality: 101.50   Quality: 101.50

Right column:

Percent Similarity: 51.163   Percent Identity: 31.008
Ratio: 1.538   Gaps: 4

Align seq 1/1 to: AL037979   from: 1 to 673

```
379 GluAsnProAsnLysGlyIleValGluProTyrSerValGluAlaTyrAs 395
          ||||||||||
130 GAAAACCCAATGAAAGGCCTCGAAAAAGCCCAAGACAGCAGAAGCA..... 174

395 nAspPheGluGluPheSerValLeuThrThrGlnAsnTyrAlaLysPheT 412
          :::||||||          |||
174 .........GACACCTCCAGTGACTAGCAAAGAAAAGCAAAAGAAGAGTATTC. 216

412 yrTyrAlaLysAsnGlySerSerGlnValValTyrCysPheAsn 428
          :::||:|||
216 ........AGAAAAGAGATGTCCCAGTTCATCGTACGAGTGCCTGAAC 255

429 AlaAspLeuLysSerProAspProAspSerGluAspGlyGlyLysThrMetTh 445
          |||||||||
256 CCTTACCGGAAA......... CCTGACTGCAAAGTGGGAAGAATTACCACAC 299

445 rProAspPheThrThrGlyGluValValLysTyrThrHisIleAlaGlyArgA 462
          | ||||||
300 TGAAGACTTTAAACATCTGGCTCGCAAGCTGACTCACGGTGTTATGAATA 349

462 spLeuPheLysTyrThrValLysProArgAspThrAspProAspThrPhe 478
          :::||||||
350 AGGAGCTGAAGTACTGTTAAGAAATCCTGGAGACCTGGAGTGCAATGAGAAT 399

479 LeuLysHisIleLeuLysLysValIleGluLysGlyTyrArgGluLysGlyGl 495
          ::|||||||
400 GTGAAACACAAAACCAAGGAGTACAATTAAGAAGTACATGCAGAGAAGTTGG 449

495 nAlaIleGluTyrSerGlyLeuThrGluThrGlnLeu 507
          |||:|||
450 GGCTGTT...TACAAACCCAAAGAGGACACTGAATTA 483
```

seq_name: gb_est78:BE744259

seq_documentation_block:

LOCUS	BE744259 797 bp mRNA EST 15-SEP-2000
DEFINITION	601576452F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3837373 5',
	mRNA sequence.
ACCESSION	BE744259
VERSION	BE744259.1 GI:10158251
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 797)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.
	Email: Robert_Strausberg@nih.gov
	Tel: (301) 496-1550
	Tissue Procurement: DCTD/DTP
	cDNA Library Preparation: Ling Hong/Rubin Laboratory
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Incyte Genomics, Inc.
	Clone distribution: MGC clone distribution information can be
	found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
	Plate: LLCM522 row: h column: 14
	High quality sequence stop: 744.
FEATURES	Location/Qualifiers
source	1..797
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:3837373"

```
BASE COUNT      241 a    180 c    197 g    176 t    3 others
ORIGIN
```

```
alignment_scores:
        Quality:       101.50          Length:       129
        Ratio:           1.538         Gaps:           4
Percent Similarity:    51.163      Percent Identity:   31.008
```

```
alignment_block:
US-09-494-297-2 x BE744259

Align seg 1/1  to: BE744259  from: 1  to: 797

379 GluAsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAs 395
    ||||||||||||     ||       ||||     |||   ||||||
 30 GAAAACCCCATGAAGGCCTCGAAAAAGGCCAGACACAGCAGAAGCA........ 74

395 nAspPheGluGluPheSerValLeuThrThrGlnAsnThrArgAlaLysPheT 412
    |||    ||||||||||:::||  |:||||||||:||::::
 75 ......GACACCTCCAGTGAACTAGCAAAAGAAAAAGAGTATTC. 116

412 yrTyrArgAlaLysAsnLysSerSerSerGlnValValTyrCysPheAsn 428
    :::||:|::     ||||||||:::|||||   ||||||:::|||
117 .........AGAAAAGAGATGTCCCAGTTCATCGTCCAGTGCCTGAAC 155

429 AlaAspLeuLysSerProProAspSerGluAspGlyGlyLysThrMeth 445
    ||||||     |||     ||||||:::|||||||||:||||
156 CCTTACCGGAAA........CCTGACTGCAAAGTGGGAAGAAATTACCACAC 199

445 rProAspPheThrThrGlyGluValLysThrTyrThrHisIleAlaGlyArg 462
    |    ||||||     |||||||   ||||||   |||:::
200 TGAAGACTTTAAACATCTGGCTCGCAAGCTGACTGACTCACGGTGTTATGAATA 249

462 spLeuPheLysThrArgValLysProArgAspThrArgAspProAspThrPhe 478
    :::|||||     |||||     |:::|||:||:|||::||||:::|||
250 AGGAGCTGAAGTACTGTAAGAATCCTGAGGAGCTGCAATGAGAAT 299

479 LeuLysHisIleIleLysLysValIleGluLysGlyTyrArgGluLysGlyGl 495
    :::|||||||     ||||     |||:||:::||||||:::|||:::|||
300 GTGAAACACAAAAACCAAGGAGTACATTAAGAGTACATCAGAAGGTTGG 349

495 nAlaIleGluTyrSerGlyLeuThrGlyThrGlnLeu 507
    |||:::|||     ||:::|||:::||||
350 GGCTGTT...TACAAACCCAAGAGGACACTGAATTA 383

seq_name: gb_est37:AV400861
```

```
seq_documentation_block:
LOCUS       AV400861      796 bp    mRNA                    EST       05-FEB-2000
DEFINITION  AV400861 Bombyx mori brain Daizo P0 (just after pupation) Bombyx
            mori cDNA clone br--1899 T3, mRNA sequence.
ACCESSION   AV400861
VERSION     AV400861.1       GI:6904513
KEYWORDS    EST.
SOURCE      domestic silkworm.
  ORGANISM  Bombyx mori
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
            ; Bombycoidea; Bombycidae; Bombyx.
REFERENCE   1  (bases 1 to 796)
  AUTHORS   Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
```

```
TITLE     Bombyx mori cDNA
JOURNAL   Unpublished (2000)
COMMENT   Contact: Mita K
          Genome Research Group
          National Institute of Radiological Sciences
          Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
          Email: kmita@nirs.go.jp
          method:uni-directional, sequence direction:sequenced from T3 primer
          (5' -> 3')
          Project='Silkworm Genome Program in MAFF, and Research for the
          Future Program in JSPS', see 'SilkBase'
          <http://www.ab.a.u-tokyo.ac.jp/silkbase/>, for whole ESTdb.
FEATURES             Location/Qualifiers
     source          1..796
                     /organism="Bombyx mori"
                     /strain="Daizo"
                     /db_xref="taxon:7091"
                     /clone="br--1899"
                     /clone_lib="Bombyx mori brain Daizo P0 (just after
                     pupation)"
                     /sex="female/male mixed"
                     /tissue_type="brain"
                     /dev_stage="P0 (just after pupation)"
BASE COUNT      332 a    123 c    174 g    167 t
ORIGIN
```

```
alignment_scores:
        Quality:       101.00          Length:       244
        Ratio:           0.835         Gaps:          10
Percent Similarity:    49.590      Percent Identity:   20.902
```

```
alignment_block:
US-09-494-297-2 x AV400861

Align seg 1/1  to: AV400861  from: 1  to: 796

473 ThrAspProAspThrPheLeuLysHisIleIleLysLysValIleGluLysGl 489
    ||||:||   |:::||||||||||||     ||||:::||||:::
 16 ACTGAAACATGATTCCTTAAAATTCTATAATTAACAAAAATT.......... 54

489 yTyrArgGluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGlyThrG 506
    :::||:::||||||   |||:::|||:||:::||||||||||||||:::
 55 ..........CGAGCTGTTTACAAAAATGGCTCAACGGCCACGA 88

506 lnLeuArg.AlaAlaThrGlnLeuAlaIleThrTyrArgPheThrAspSerAr 522
    |||:::||||||   |||||:::|||||||||||||||||:::
 89 AGATCAAACGCCTGGAAAATCGTATGGCT........AAACTCCTTGACGAGGA 132

522 aGlyLeuAspLysAspLysLysAspPtyrHisGlyPheGlyAspMetA 539
    |||     |||:::|||||||||     |||||||||||||:::
133 GGAAGAGGATGAATTTTACAAAACAACATACGGAGGGTTTAATGATGTTG 182

539 snAspSerThrLeuValAlaAlaLysIleLeuValGluTyrAlaGlnAsp 555
    :::|||||||||||||||||||||     ||:::|||||||||:::
183 AAGAGGAC........GTAGAAAATCAACACAGAA 208

556 SerAsnProProGlnLeuThrAspLeuAspPhePheIleProAsnAsnAs 572
    :::|||     |:::|||||||||||||     ||||||::::
209 AAGGAAGTTCATGATGAAGCTTGATTCAGATTTTGATATTGCATGAGAATGA 258

572 n............................LysTyrGlnSerLeuIleG 579
    |||     :::|:::
259 TGAACCAGTGTCTGACCAGGAGAGTTGAAGAAAAAAGCTAAAAGAAAAAGTTG 308

579 lyThrGlnIrpHisProGlu. 585
    |||||:::     :::
309 GCACTAAAGCTTATAAGGAGCCCCAATAAGAAGAAAATTAAATGAAAAATCA 358

586 ...........................AspLeuValAspIl 590
    :::||     |||
359 AAAAAGACAGTCATCAAAAAAATCCAAAGATGATAAATCTATAGAAGACA 408
```

seq_documentation_block:

```
LOCUS       BE795700                1056 bp    mRNA            EST       20-SEP-2000
DEFINITION  601590702F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3944761 5',
            mRNA sequence.
ACCESSION   BE795700
VERSION     BE795700.1  GI:10216898
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM802 row: c column: 02
            High quality sequence stop: 724.
FEATURES             Location/Qualifiers
     source          1..1056
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3944761"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      331 a    242 c    306 g    177 t
ORIGIN
```

seq_name: gb_est78:BE795700

```
590 etIleArgMetGluAspLysLysGluValIleProValThrHisAsnLeuT 607
    |||||:||:
409 AGTTGAAAAACAAAACAAAAGTGCCGTGGACAAAAATTGTTGATACCTCAA 458

607 hrLeuArgLysThrVal.........ThrGlyLeuAlaGlyAspArgThr 620
    |||||:::|||                  ||::|||||
459 TAGAACGGAAGTCTATAAGACAAAAGTACAGCTATGAAATCTCCGAGACA 508

621 ...LysAspPheHisPheGluIleGluLeuLysAsnAsnLys.......... 633
    :::         |:      |: |  ::::|   ::
509 ATACAAAAGGATAAAAATCAGGTCAGAGCTGAAAAAGAAGAAAACCGAAGA 558

634 ................GlnGluLeuLeuSerGlnThrThrValL 642
                    ||||||||||:||::|  |||
559 AGTAGAAGAGAAGAGATGCCGACCCAGGAAGAGCTGTTGAAGAAGCAGTCG 608

642 ysThrAspLysThrAsnLeuGluPheLysAspGlyLysAlaThrIleAsn 658
    :::||::|
609 AGACAGAAGAAAGAAGAACTTAAAG...................     634

659 LeuLysHisGlyLeuSerLeuThrLeuGln 668
    :::||::|       |||::
635 TTAGAACGTTTGAACAAATCGAACTAGAA 664
```

alignment_scores:

```
        Quality:        101.00          Length:        356
        Ratio:          0.656           Gaps:          21
Percent Similarity: 43.258      Percent Identity: 22.753
```

alignment_block:
US-09-494-297-2 x BE795700

Align seg 1/1 to: BE795700   from: 1 to: 1056

US-09-494-297-2 from: 1 to: 1056

```
206 GluSerGluSerAsnLeuValSerThrSerGlnLeuSerLeuMetArgGl 222
    |||         :::|:  |  |::::|  ||:      |::||
11  GAAATGACTNCGAACCGGCTTGCAGAGAGCCTTCGCTTTGAGCCAACA 60

222 nAlaLeuLysGlnLeuIleLeuAspProAsnLeuAlaThrLysMetProLysG 239
    :::|||  || |  :::    ||||||| |||||::   |  ||::
61  G......GAAGAACTAGCGGAT.......                     76

239 lnValProAlaAspAspPheGlnIleLeuSerIlePheGluSerGluAlaLysGly 255
    ||:|    |||  | ||||||||||::||||||||:|||
77  ..TTCCCAAAAGACTACCTCTTGAGT......GAGAGTGAAGATGAGGG 118

256 AspLysTyrTyrAsnLysGly...........TyrGlnAsnLeuLeuSerGlyGl 269
    :::|||  |       :::||:::||||||||
119 GAC......AATGATGGAGAGAAGAAAAGCATCAAAAGCTTCTG 154

269 yLeuValProThrLysProProThrProGlyAspProProMetProProA 286
    :::
154 .....                                            154

286 snGlnProGlnThrThrSerValLeuIleArgLysTyrAlaIleGlyAsp 302
    :::|:|  |||   ||   ||||||||  ::|
235 CAATGTC......GAAGCAATCAGTTCCCTGATGGAAAG 184

303 TyrSerLysLeuLeuGlu.............GlyAlaThrLeuGlnLeuThrGlyAs 317
    |||||::|  |||      |:::::||||
185 AGGGCGGAAAATTGGCTCGAGAGGTCTTGAGGGTGTCGCAGAGTT 234

317 pAsnValAsnSerPheGlnAlaArgValPheSerSerAsnAspIleGlyG 334
    |::||:::   :::|||||||||  :::
235 CAATGTC......                                     260

334 luArgIleLeuLeuSerAspGlyThrTyrThrTyrThrLeuThrGluAsnSer 350
    ||||:::      |:::|||
261 AAAAACCTGGTCCTTGCCAGAT......                      280

351 ProAlaGlyTyrSerIleAlaGlyGluProIle........ 360
    |||||||:::
281 ......CTGCTTGAGCCTGTTAAAACTTCAnCTCTTGGC 315

361 .ThrPheLysValGlnLeuGlyLysValTyrThrIleIleAlaAspGlyGlyLysG 377
    |||    |::     ||  |::::||     |:
316 CACTGTGAAAAAGCAACTCAGTCAGAGTCAAATCA.......AAGAAGA 356

377 lnIleGluGluLysAsnPro......AsnLysGluIleGluGluGluProTyrSerValGlu 392
    |||                         ||  ||    |::  |||
357 CAGTGGGAGTTACCTCTGAACAAAGAGAGATTGAACGGATCCACAGAGAA 406

393 ...AlaThrTyrAsnAspPheGluPheSerValLeuProThrThrGlnAsnTy 408
    :::|||||||||       ||:::
407 GTACCATTCAATAAA......ACCGACAAAGTCCT 435

408 rAlaLysPhe.......TyrTyrTyrAlaLysAsnLeuLysAsnGlySerSerGlnV 423
    ::::|||  :::   |||||:::  | ||    :::
436 CTCCAAATGGGACCCTGTCGTCCTGAGAACCGGCAG......GCAGAGAGC 482

423 aValAlaTyrCysPheAsnAlaAspLeuLysSerProProAspSerGlnLeuAsp 439
    |:::|||||||         |::  |::|:|||
483 TGGTTTTTT.......CCCCTGGAAAGAGAGGAG 508

440 ..........GlyGlyLysThrMetThr 445
    |::||:::
509 CCAGCCCATTGCTCCCATTGAACAnTGTGCTCAGTGGCTGGAAGGCAAGAAC 558
```

```
445 rProAspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgA  462
                                              :::
559 TCCC.............................CTGGAGCAGG  572

462 spLeuPheLysTyrThrValLysProArgAspThrAspThrPhe  478
    ::::::||||::|                           ||:::
573 AAATTTTCAACCTCCTCCATTAAGAACAAGCAGCCAGTGACACAGCCCTTTA  622

479 LeuLysHisIleLeuLysLysValIleGluLysGlyGlyTyrArgGluLysGlyGl  495
    |||   :::  :::::                         TCTCTCCGAGCCAT  657
623 CTGACCCCTGTGGAAAAGGCC.............

529 LeuLysAspTyrHisGly  534
    ::: ::: :||:
758 GTAACAAGTTCCCAAGGC  775

512 lnLeuAlaIleThrTyrTyrPheThrAspSerAlaLeuGluLeuAspLysLysAspLys  528
    |||||||                               :::|||  |||
708 CTCTGCAGTCCTCCTTACTATGACGGCCAAGGCTCGAAGAGAGAGAGAGAAAATCAA  757

495 nAlaIleGluTyrSerGlyLeuThrGlnLeuArgAlaAlaThrG  512
    ::::|||||||::|       :::|||||:::          |||
658 GAGCCTAGAAGAGGCAAAAGATGCGAGCAGCAGGAGCTTCAGAGAGGGCTCGGG  707

seq_documentation_block:
seq_name: gb_est90:BF684247

LOCUS       BF684247     946 bp    mRNA       EST       22-DEC-2000
DEFINITION  602141162F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4302470 5',
            mRNA sequence.
ACCESSION   BF684247
VERSION     BF684247.1  GI:11969655
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 946)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1164 row: k  column: 15
            High quality sequence stop: 707.
FEATURES             Location/Qualifiers
     source          1..946
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4302470"
                     /clone_lib="NIH_MGC_46"
                     /tissue_type="leiomyosarcoma cell line"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT    279 a    212 c    277 g    178 t
ORIGIN
```

alignment_scores:
```
Quality:              99.50            Length:      339
Ratio:                0.578            Gaps:        19
Percent Similarity:   50.737          Percent Identity: 23.304
```

alignment_block:
us-09-494-297-2 x BF684247

Align seq 1/1   to: BF684247   from: 1   to: 946

```
422 GlnValValIleTyrCysPheAsnAlaAspLeuLysSerProProAspSerGl  438
    |||                                :::
37  CAAAATGTCTACTGT...............TCACGAAATCCTGTGCAAGCTCA  74

438 uAspGlyGlyLysThrMet.ThrProAspPheThrThrGlyGluValLys  454
    |||||||        :::|||  :::          |||:::||||||||
75  GCTTGGAGGGGTGATCACTCTACACCCCAAGTGCATATGGGTCTGTCAAA  124

455 ..TyrThrHisIle...AlaGlyValArgAspLeu........PheLysTyrTh  467
      |||||||::::::     |||  |||||||             |||  :::
125 GCCTATACTAACTTTGATCCTGAGCGGGATGCTTTGAACATTGAAACAGC  174

467 rValLysProArgAspThrAspThrPhe.........LeuLysHisI  482
    ::::||                      |||:::            ::|
175 CATCAAGACCAAAGGTGGATGAGGTCACCATGTCAACATTTGACCAAC  224

482 leTyrLysLysValIleGluLysGlyGlyTyr.....................Arg  491
    ::::|||                         |||:::
313 CTGGCCACCTGGACGGTGATTTGGGCCCTATGAAGACACCTGCTCAGTA  362

492 GluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGlnLeuArg  508
    |||  :::          ::|||||||
225 CGCAGCAATGCACAGAGACAGGATATTGCCTTCGCCTACCAGAGAAGGAC  274

508 rgAlaAlaThrGlnLeuAlaIleThrTyrTyrPheThrAspSer.AlaGluLeu  524
    |||  :::          |||||||   :::          |||
524 uAspLysLysAspLys...AspTyrHisGlyPheGlyAspMetAsnA  540
    |||          :::           ||||||        :::
363 TGACGGCTTCTGAGCTAAAAGCTTCCCATGAAGGGGCGCTGGGAACCGACGCAGG  412

540 spSerThrLeuAlaValAlaLysIleIleLeuValGluTyrArgAlaGlnAspSer  556
    |||||                             |||  :::
413 ACTCT...........................CTCATTGAGATCANCATCTGCCCAGA  441

557 AsnProProGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLy  573
    :::|||                               ||:::
557 AsnProProGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLy  573
442 ACCAACCAGGAGCTGCCAGAAATT.................AACAGAGT  473

573 sTyrGlnSerLeuIleGlyThrGlnTrpHisProGluAspLeuValAspI  590
    :::|||::::::             |||
474 CTACAAGGAGAAAATGTACAAGACT.................GATCTG  501

590 leIleArgMetGluAspLysLysGlyLeuValValIleProValAlaThrHisAsnLeu  606
    :::       ||||||::::::||            |||:::
502 GACTTCCGCAAGCTGCATGGTTGGCCTGGCAAAGGGTAGAAGACCAAGAGGA  531

607 ThrLeuArgLysThrValThrGlyValLeuAla...GlyAspArgThrLysAs  622
    :::|||              |||::::::        ||||||::::::
502 ..................GAGAAGGACACATTATTTCGGACACACATCTGGT  531

622 pPheHis...PheGluIleGluLeuLysAsnAsnLysGlnGluLeuLeuS  638
    :::|||    ||||||::::::||              |||:::
622 GACTTCCGCAAGCTGCATGGTTGGCCTGGCAAAGGGTAGAAGACCAAGAGGA  622

638 erGlnThrValLysThrAspAspAlaGluProLysAspAspGlyLys  654
    |||:::|||::::::     |||:::          |||:::
582 TGGCTCTGTCATTGATTATGAACTGATAGACCAACGATGGGTCCGGGATCT  631

655 Ala.................................  655
632 CTATGACGTTGGAGTGACGAGGAAAAGGAACTGATGTCCCAAGTGGGATCA  681
```

alignment_block:
US-09-494-297-2 x AU139155

alignment_scores:
  Quality: 99.00   Length: 272
  Ratio: 0.780   Gaps: 16
Percent Similarity: 46.691   Percent Identity: 24.632

seq_name: gb_est30:AU139155

seq_documentation_block:

```
LOCUS       AU139155           800 bp    mRNA                    EST       25-OCT-2000
DEFINITION  AU139155 PLACE1 Homo sapiens cDNA clone PLACE1010047 5', mRNA
            sequence.
ACCESSION   AU139155
VERSION     AU139155.1  GI:11000676
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 800)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..800
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1010047"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT      212 a    227 c    184 g    173 t    4 others
ORIGIN
```

682 GCATCAТGACGAGCGGAGCGGAGCGGTGCCACCTCCAGAAAGTTATCGATAGGTAC 731
    |||
656 ....ThrIleAsnLeuLysHisGlyGluSerLeuThrLeuGlnGlyL 670

732 AAGAGTCACAGGCCCTTATGAAAATGGGAAGCAATCGGAGAGTAAAAGGAG 781
    ||| |||| |||:::::
670 euProGluGlyTyrSerTyrLeuValLysGluThrAspSerGluGlyTyr 686

782 AGCCGAAAATGGTTCTGGACCGGGTCAGGGACTAGAACAGGCGGGATTG 831
    |||||||| ::: |||
687 LysValLysValAsnSerGlnGluValAlaAlaAsnAlaThrValSerLysTh 703

832 GCGAACGGGGAGGAGAAACCAAAGGGGAACAGAGAATAAGGCGACAC 881
    ::::::|||||||||||| ::: |||:::
703 rGlyIleThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProV 720

882 ACGGGTCCCAGGGAGGACTGGGAAAATCGGCAACAGAGAGAGAGGAGCCGG 931
    ::::   :::::::::::||||||:
720 alValProThr 723

932 CAAATCCAACC 942
    |||||||
                942

Align seg 1/1 to: AU139155   from: 1   to: 800

503 ThrGlyThrGlnLeuArgAlaAlaThrGlnLeuAlaIleTyrTyrPheTh 519
    ||| |||::: |||:::: ||||||| ::::|
70 ACCATCACAGGTTTACAACCAGGCACTGACTACAAGACACTACCTGTACAC 119

519 rAspSerAlaGluLeuAspLysAspLysAspLysAspTyrHisGlyPheG 536
    ....
120 C... 120

536 lyAspMetAsnAspSerThrLeuAlaValAlaValAlaIleLeuValGluTyr 552
    ::::|||||||||| :::   :::::::||||||:
121 .......TTGAATGACAATGCT.....TCCCCTGTGGTCATCGAGCGC 162

553 AlaGlnAspSerAsnProGlnLeuThrAspLeuAspPhePheIlePr 569
    ::::::|||||||| ||| ||| |||
163 TCCACTGCCATTGATGCACCA.......TCCAACCTTGCGTTTCCTGGCCAC 206

569 oAsnAsnAsnLysTyrGlnSerLeuIleGlyThrGlnThrHisPro... 584
    ||||||||| :::|||:||:||:
207 CACACCCAAT.........TCCTTGCTG...GTATCATGGCAGCCGCCAC 244

585 ....GluAspLeuValAspIleIleLeuArgMetGlu........... 595
    :: ||||||: |||
245 GTGCCAGGATTACCGGGCTACATCATCAAGTATGAGAAGACCCTGGGTCTCCT 294

596 LysLysGluValIleProValThrHisAsnLeuThrLeuArgLysThrVa 612
    :::||||||||||||::::   :::
295 CCCAGAGAAGTGGTCCCTCGGCCCCGGCCCTGGTGGTCACAGAGGCTACTAT 344

612 lThrGlyLeuAlaGlyAspArgThrLysAspPheHisPheGluIleGluL 629
    :|||||||||| ||| :
345 TACTGGCCCTGGAA.....CCCGGGACCGAATAATACAATTATGTCATTGCCC 391

629 euLysLysAsnAsnLys........................... 633
    ||||||||||||||||:::
392 TGAAGAATAATCAGAAGAGGCGAGCCCCTGATTGGAAGGAAAAAGACAGGA 441

634 GlnGluLeuLeuSerGlnThr........................ 640
    |||||| ||||||||||
442 CAAGAAGCTCTCTCTCAGACAACCATCCATGGCCCCCATTCCAGGACAC 491

641 ....ValLysThrAspLysThrAsnL 648
                    |||||||:::
492 TTCTGAGTACATCATTTCATGTCAATCCTGTTGGCCACTGATGAGAACCCT 541

648 euGluProLys.....AspGlyLysAlaThrIleAsnLeuLysHisGlyGlu 663
    |||:::||||::: ||| :::||||
542 TACAGTTCAGGGTTCCTGGAACTTCTACC............... 570

664 SerLeuThrLeuGlnGlyLeuProGluGlyTyrSerTyr...LeuValVal 679
    ||| ||||| ||| ||| :::: |||:::
571 AGTGCCACTTCTGACAGGCCCTCACCAGAGGTGCCACCTACAACATCATAGT 620

679 sGluThrAspSerGluGlyTyrLysValLysValAsnSerGlnGluValAla 696
    :::   ::::::: ||||||||::::::::: |||:
621 GGAGGCACTGAAAGACCAGCAGGCATAAGGTTCGGGAAGAGGTTGTTA 670

696 laAsn.AlaThrValSerLys.......Thr GlyIleThrSerAspI 709
    :: ||||||||||| |||:::   :::|||:
671 CCGTGGGCAACTCTGTCAACGAAGGCTTGAACCAAACCTACGGGNTGACTC 720

709 uThrLeuAlaPheGluAsnAsnLysGluProValValProThrGlyValAla 726
    :: |||::|: ||||||||||| |||
721 G.....TGCTTTGACCCCTACACAGTTTCCATTATGCCGTTTGGAGANG 764

726 spGlnLysIleAsn 730
    || ::: |||
765 ATGGGGAAANGAAT 778

seq_name: gb_gss23:AZ535645

seq_documentation_block:

```
LOCUS       AZ535645        875 bp    DNA        GSS        03-NOV-2000
DEFINITION  ENWTDK49TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ535645
VERSION     AZ535645.1  GI:11092592
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica.
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            High quality sequence start: 22
            High quality sequence stop: 749.
            Class: shotgun
FEATURES             Location/Qualifiers
     source          1..875
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                     method for isolate identification. Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + 1 method used for
                     the library construction is described in detail in Smith,
                     H.O. and Venter, J.C. (Making small insert libraries for
                     whole genome shotgun sequencing projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT    516 a    80 c    144 g    135 t
ORIGIN
```

```
alignment_block:
US-09-494-297-2 x AZ535645

alignment_scores:
    Quality:              99.00
    Ratio:                0.723
Percent Similarity:      49.104

Align seg 1/1   to: AZ535645      ..

Align seg 1/1   to: AZ535645    from: 1  to: 875
                                               Length:          279
Percent Similarity:  49.104   Percent Identity:  21.864            Gaps:             16
```

```
483 LysLysValIleGluLys..........GlyTyrArgGluLysGlyGlnAl 496
    |||||||||||:::::::..........::||||:||||||||||| :::
 29 AAAAAGGTAATTGAAAAGACCAAAGAAGAGAGTTTAAGAAAAAACAAAAGA 78

496 aIleGluTyrSerGlyLeuThrGlnThrGlnLeuArgAlaAlaThrGlnL 513
    |||:||||||||||||||||||||||:|:::||||||||||||||||||||
 79 TTGGGAG.......... GAAAAAGACATTAAGGCTATA......... 106

513 euAlaIleTyrTyrPheThrAspSerAlaGluLeuAspLysAspLysLeu 529
    :::||||::||||||||||||:||||||::||:|||||||||||::::::
107 ................................GAGCAGAAATT 118

530 LysAspTyrHisGlyPheGlyAspMetAsnAspSerThrLeuAlaValAlaI 546
    |||:::::||||||||||||||:|||||||||||||||||||||||||:||
```

```
seq_name: gb_est81:BE970296

seq_documentation_block:
LOCUS       BE970296        862 bp    mRNA        EST        04-OCT-2000
DEFINITION  601794531R1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949908 3',
            mRNA sequence.
ACCESSION   BE970296
VERSION     BE970296.1  GI:10583229
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   NIH_MGC http://mgc.nci.nih.gov/,
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
119 AAAAACAAAAACCCAATTAAAGAAATTGAATGAT..AGATGTAAAATTGC 165
    |||:::     :::     :::::::||||||||||     ::||
166 AAGAATAATA.................................. 175

546 aLysIleLeuValGlyThrValAlaGlnAspSerAsnProProGlnLeuThrA 563
    |::||||||     |||:::||||||||||||||||||||||||||||
175 ...........................................175

563 spLeuAspPhePheIleProAsnAsnLysTyrGlnSerLeuIleGly 579
    |||||||||||     |||||||||||||||||||:::::::
176 .....................AATAACAAA..CAAAGTATAATGAAA 199

580 ThrGlnThrProHisProGluAspValAspIleIleAlaArgMetGlu...As 595
    ||||||||     |||::::     :::|||     ||
200 ACACAAGAAGAGAAACAGAAAAAAATAAAGAAGAAGAAGAAGAAGAAAAAGA 249

595 pLysLysGluValIleProValThrHisAsnLeuThr...LeuArgLysT 611
    ||||||||     ||||     ||||||||||||||||::||     |||
250 TAAAGAAGAAATACTTGCTCTGAAAAATCAAATTACTGAATTAAAAGAAA 299

611 hrValThrGlyLeuAlaGlyAspArg.................. 619
    |:::     |||||||||||||::::|||
300 AACTTAAAGAATTCAGCATTGGAAGAATGGAAGAAGAACAAGCCATCTCA 349

625 eGluLeuIle...................GluLeuLysAsnAsnLysG 634
    |||||:     ...................|||||||||:::|||||
350 GAAGAAATGATGAGAAAAAGCAAAAGAAATAGTTAGGAAGAAATTTGAAGA 399

634 InGluLeuLeuSerGlnThrValLysThrAspLysThrAsnLeuGluPhe 650
    ::||||:::     :::|||||||||||||||||||||||||||:::
400 AGAAAATTACAGAGAAAATGAAAAACACAAAATTAAAAAGATTCAAACGAATTATG 449

651 LysAspGlyLysAlaThrIleAsnLeuHisGlyGluSerLeuThr... 666
    :::     |||||||||||||||     |||||||||||||||:||
450 AAGAAATGAAGGAAAAGAGAAAAATGAAAAATTAGAGAAGAAGAAAAACTTAAAACTA 499

667 ..................LeuGlnGlyLeuProGluGlyTyrTyrSerT 676
    ..................||||||||||||||||||||||||:|||
500 CAA......GGAAAA......ATAAATGAGATAGGAAGGAGAAAAAATCACAGA 540

676 yrLeuValLysGluThrAspSerGluGlyTyrArgLysValLysAsnSer 692
    :::||||||||||||     ||||||||||||     |||
541 AGTAAATAATTAAAGAGAAGAAAAAATCCGAGCATCCAAGTAATAAAAAGA 590

693 GlnGluValAlaAsnAlaThrValSerLysThrGlyIleThrSerAsp.. 708
    |||     :::     |||     |||::|||||||::::::|||
591 AAATTGAAAAAGAA..........AATGAAGAAAATGAAGAAGAAAAATTGAAAAG 634

635 CTAAAAAAGAGAAAACAAAAACGCTAGAACAAAAATGCAAAACACTAGAGA 684
    |||::     ||||||||     :::::::||||||     ::::
685 GAAAAATTGAAACGATAGAG...GAAAAATACTAAAGA 718

709 .......GluThrLeuAlaPheGluGluLeuAsnAsnLysGlu 718
    .......::||||||||     |||     :::
635 CTAAAAAAGAGAAAACAAAAACGCTAGAACAAAAATGCAAAACACTAGAGA 684
```

Left column:

FEATURES
  source
      Tel: (301) 496-1550
      Email: Robert_Strausberg@nlh.gov
      Tissue Procurement: CLONETECH Laboratories, Inc.
      cDNA Library Preparation: CLONETECH Laboratories, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLCMB15 row: 1 column: 13.
  1. .862
      Location/Qualifiers
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:3949908"
      /clone_lib="NIH_MGC_78"
      /lab_host="DH10B (T1 phage-resistant)"
      /note="Organ: pancreas; Vector: pDNR-LIB (Clontech);
      Site_1: SfiI (ggccgcctggcc); Site_2: SfiI (ggccattatggcc
      ); 5' and 3' adaptors were used in cloning as follows: 5'
      adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
      sequence: 5'-ATTCTAGAGCGCGGCGGGCCGGACATG-dT(30)BN-3',
      (where B = A, C, or G and N = A, C, G, or T). Average
      insert size 1.2 kb (range 0.5-4.0 kb). 14/15 colonies
      contained inserts by PCR. This library was enriched for
      full-length clones and was constructed by Clontech
      Laboratories (Palo Alto, CA)."
BASE COUNT   147 a   243 c   256 g   216 t
ORIGIN

alignment_scores:
  Quality: 98.50
  Ratio: 0.912
Percent Similarity: 46.352   Percent Identity: 22.747

alignment_block:
US-09-494-297-2 x BE970296/rev

Align seg 1/1 to reverse of: BE970296   from: 1 to: 862

              Length: 233
              Gaps: 11

112 LysLysAlaPheProLeu................GlySerAspSerSerVa 123
    ||:||||::||:||:||| ||:|||:||
610 CAAAAAGCGTTGCCCAGGGGCCCCCAAACGGGAGGGGCCCCCCCAGGGGGT 561

123 lLysLysTrpTyrLysLysHisAspGlyIleSerThrLys........... 136
    ||||||| ||| :::::::||
560 CACCAAATGGATGAAAAAATCCCGGGCGGAAAAAAACAAGGAGGGGCCCC 511

137 ...PheGluAspTyrAlaMetSerProArgIleThrGlyAsp 149
       ||| :::   :||
510 CCAGGAAATTGTTTCCTGAACCAGGGCGGGAAGCCC........... 476

150 GluLeuAsnGlnLysLeuArgAlaValAlaMetThrValArgHisProGlnAs 166
      :::   :::::::||
475 ......CAGAGTTCGGCCCCCCGCGAGATACAAACCACCGGCGCCAA 435

166 nAlaAsnGlyIleMetGluGlyLeuGluProLeuAsnAlaIleArgValT 183
                   :||:|::|||
434 C.................................CTTGTCAGATTAAGGCTA 415

183 hrGlnGluAla...ValTrpTyrTyrSerAspAsnAlaProIleSerAsn 198
    ||||:| ||| :::::::||:|:|
414 CACAGAAAACAAAGACACAGGTTTCTTCTCCGGAATAATAAGGACGCCGCAC 365

199 ProAspGluSerPheLysArgGluSerGlnSerAsnLeuValSerThrSe 215
       :|| || |:||:||||
364 CACGAAGGAGTTTC.................AACGACCCCTTTCGAACCC 330

215 rGlnLeuSerLeuMetArgGlnAlaLeuLysGlnLeuIleAspPro.... 230
      |:| |||| :::::||
329 GGGGATTGCGACTCATCCCCGAGGGGATT......CTGGCGCAACCATCAA 286

Right column:

231 ....AsnLeuAlaThrLysMetPro 237
                     |||::: ||:||:|||
285 AGAAATGTGGGAGGAAGCCAACACAGGTAAAACATAAGGAGGAATCAACCA 236

238 LysGlnVal.ProAspAspPheGlnLeuSerIlePheGluSerGluAspL 254
    :::::||||:||| :||||| |||:||:::
235 AGGGACATCCCC........CGGAGTCTCCATGGCACAGAGGAGAGC 198

254 ysGlyAspLysTyr............AsnLysGlyTyrGlnAsnLeuLeu 266
                          :|:|||
197 GGACCCACCAATTCTTAGTGGGGCGCCCAACCATGGGGACCCACTCTCCAG 148

267 SerGlyGlyLeuValProThrLysProProThrProGlyAspProProMe 283
       :::::|||| :::|||||| ||| |||
147 AATGCAGGTACGGCGTCCGGAGAACCCCCCGGGAGGGGGGAATGCCTGC 98

283 tProProAsnGlnProGlnThrThrHrSerValLeuHisArgLysTyrAlaI 300
      ||| :::||||||| :::|||||| :::
97 TCCCGCGTCCCTACCCCGTGTGGGGAGACTCCTTTATTGGCACGACACGGG 48

300 leGlyAspTyrSerLysLeuLeuGluGlyAlaThrLeuGlnLeuThr 315
       ||| :::|| :|||:|| :|||
47 CAGGCTTCTTCAAAAAAACAACAACAAAAGGGGAATAACCCACAAAATGACA 1

seq_name: gb_est26:AI895477

FEATURES
  source   Location/Qualifiers
       1. .566
       /organism="Lycopersicon esculentum"
       /cultivar="TA496"
       /db_xref="taxon:4081"
       /clone="CLEC7N6"
       /clone_lib="tomato callus, TAMU"
       /tissue_type="callus"
       /dev_stage="25-40 days old"
       /lab_host="Xl1-Blue MRF'"
       /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
       XhoI; supplier: Giovannoni laboratory; CLEC - Cotyledons
       of seedlings 7-10 days post-germination were excised, cut
       at both ends and placed on MS medium with no selection.
       Mixed callus was harvested at 25 and 40 days and included
       undifferentiated masses. Tomato Callus EST Library"
BASE COUNT   192 a   96 c   137 g   141 t
ORIGIN

alignment_scores:

```
                 Quality:   98.00                    Length:      176
                  Ratio:     1.101                    Gaps:          9
Percent Similarity:  50.568      Percent Identity:  27.841

alignment_block:
US-09-494-297-2 x AI895477

Align seg 1/1  to: AI895477  from: 1  to: 566

 131 AspGlyIleSerThrLysPheGluAspTyrAlaMetSerProArgIleTh 147
     |:|||||||  | :||||||| ||         ||     |:     ||
 152 GATGGGATACTTAAGTCTATTTCAGCAATATAGTGGGCAATATCGGCCAC 201

 147 rGlyAspGluLeuAsnGlnLysLeuArgAlaValMetTyrAsnGlyHisP 164
     |  | |:|:|||||::||     |||         || |||  |||||
 202 TGATGATAGCCTCAACACCTTCTTATCATTTCTGAAGGAAAATGGA..... 247

 164 roGlnAsnAlaAsnGlyIleMetGluGlyIleuGluProLeuAsnAlaIle 180
                                             ::|||
 248 ..........................................GTCAACCTG... 256

 181 ArgValThrGlnGluAlaValTrpTyrTyrSerAspAsnAlaProIleSe 197
                                          |||:|:||| |||::
 257 ....................................GATGAGGCTGAGATAAA 273

 197 rAsnPro.........AspGluSerPheLys.........ArgGluSerGluS 209
     :::|:||          |||||||| ||          |:|::::|||||
 274 AAAGCCAAGAGATGATGAGAGCTATGAAGAGGGAAAAGTCATCTGAGA 323

 209 erAsnLeuValSerThrSerGlnLeuSerLeuMetArgGlnAlaLeuLys 225
     ||||| | ||:|||:|:|||||||    :::   |::::||||:::
 324 GTAAC........TCTGCGTCTGATCTTTCAAACTATATCAGAATCCCTTCAA 367

 226 GlnLeuIleAspProAsnLeuAlaThrLysMetProLysGlnValProAs 242
                                   ::||||||||::::
 368 ..........ATTGAT..................TTACCGAAGGAAGAGGAAA 393

 242 pAspPheGlnLeuSerIlePheGluSerGluAspLysGlyAspLysTyrA 259
     |||||||     :::    :::::::|:|| :
 394 AGAC........TTGTCAGGATTAGTGATAAATCCACAAGCTGAAAAGGCAA 437

 259 snLysGlyTyrGlnAsnLeuLeuSerGlyGlyTyrLeuValProThrLysPro 275
     ::::::|||::||||||||||||||||||||     :::
 438 GTAACAACTATAAGAGAACACTCTCGGGTGGTCTT..........GAGAGC 478

 276 ProThrProGlyAspProMetProProAsnGlnProGlnThrThrSe 292
                                           ::::::::
 479 CCGAGAGCTGAGGTGCCAAAAACTGCAATACTGCAAAGGATTAATTCAAA 528

 292 rValLeuIleArgLysTyrAlaIleGly 301
     |||                 ::::||||
 529 GAACTTATCAAAAGTCATATCAATTGGGT 556
```